

Survey Administration

Summary

In October 1998, the FY98 Army-wide Survey of Civilian Personnel Management was mailed to the home address of a random sample of civilian employees. A follow-up letter was mailed to all sampled personnel two weeks later. The following is a breakdown of the sample surveyed:

- 8,279 civilian employees
- 6,553 civilian supervisors
- 2,887 CPA employees
- 300 CPA supervisors

Analysis Methodology

The two data sets, corresponding to the supervisor and the employee surveys, were individually read into SPSS data sets. Each data set contained a "form code," telling whether an individual respondent was civilian or CPA. For all analyses, the data from each survey were split into the separate groups, using the SPSS "Split File" command. For the individual items, frequencies were run, using the SPSS "Frequencies" command. Then means and standard deviations were taken on the individual items, using the SPSS "Descriptives" command.

For the composites, "multiple response" groups were defined for the SPSS data set, containing all the individual items that go into the composites. The multiple response feature of SPSS treats items in a group as if they were all sub-parts of a single item; all individual item scores are tabulated independently, without regard to an individual respondent's other responses. That is to say, when six items go into a multiple response group, the valid number of cases to be tabulated in the multiple response group is six times the number of respondents; the *response* is the unit of analysis, not the *respondent*. The composite frequencies were attained by using the "Mult Response" command and "Frequencies" subcommand in SPSS.

The means and standard deviations for the composites, however, could not be attained simply from running a simple SPSS command on the data set. The means were computed by bringing the frequencies into an Excel spreadsheet, multiplying the number of responses in each category by the value applied to the category (e.g., 'strongly disagree' responses were assigned a 1), summing the total scores for each category, and dividing by the total number of responses:

$$1(n_1) + 2(n_2) + 3(n_3) + 4(n_4) + 5(n_5)$$

$$(n_1 + n_2 + n_3 + n_4 + n_5)$$

The mean score for the composite, then, is simply the value of the average response across *all* respondents, with no summary information for each respondent¹.

These means were verified by also manipulating the data such that all items included in a composite were in one variable (e.g., the responses for questions 13-17 ended up stacked on top

¹ The composite scores sent earlier were averaged *across* composite items, *within* respondent, resulting in a tabulation of respondents' average responses, rather than a tabulation of *overall* responses. This method was rejected because it was not equivalent to the methods used in previous studies and because finding means for individuals tends to pull the frequency distribution toward the middle ranges, and while it better represents individual respondents' overall impressions, it doesn't preserve the extreme responses on the individual questions.

of each other in the SPSS data editor), and then running the regular SPSS “Descriptives” command on them.

Because of the need to pool variances across the questions, this stacking method would not be a valid way to determine the standard deviation. Instead, the individual variances (standard deviation squared) were taken, multiplied by one less than the number of valid responses, summed, divided by the *total* number of valid responses *minus* the number of questions being pooled, and then the square root of this number was taken:

$$\sqrt{\frac{(s^2_1 (n_1 - 1) + s^2_2 (n_2 - 1) + s^2_3 (n_3 - 1) + s^2_4 (n_4 - 1) + s^2_5 (n_5 - 1))}{(n_1 + n_2 + n_3 + n_4 + n_5 - 5)}}$$

The standard deviations were checked for plausibility against the individual item deviations, and all numbers going into the calculations were double-checked for quality assurance.